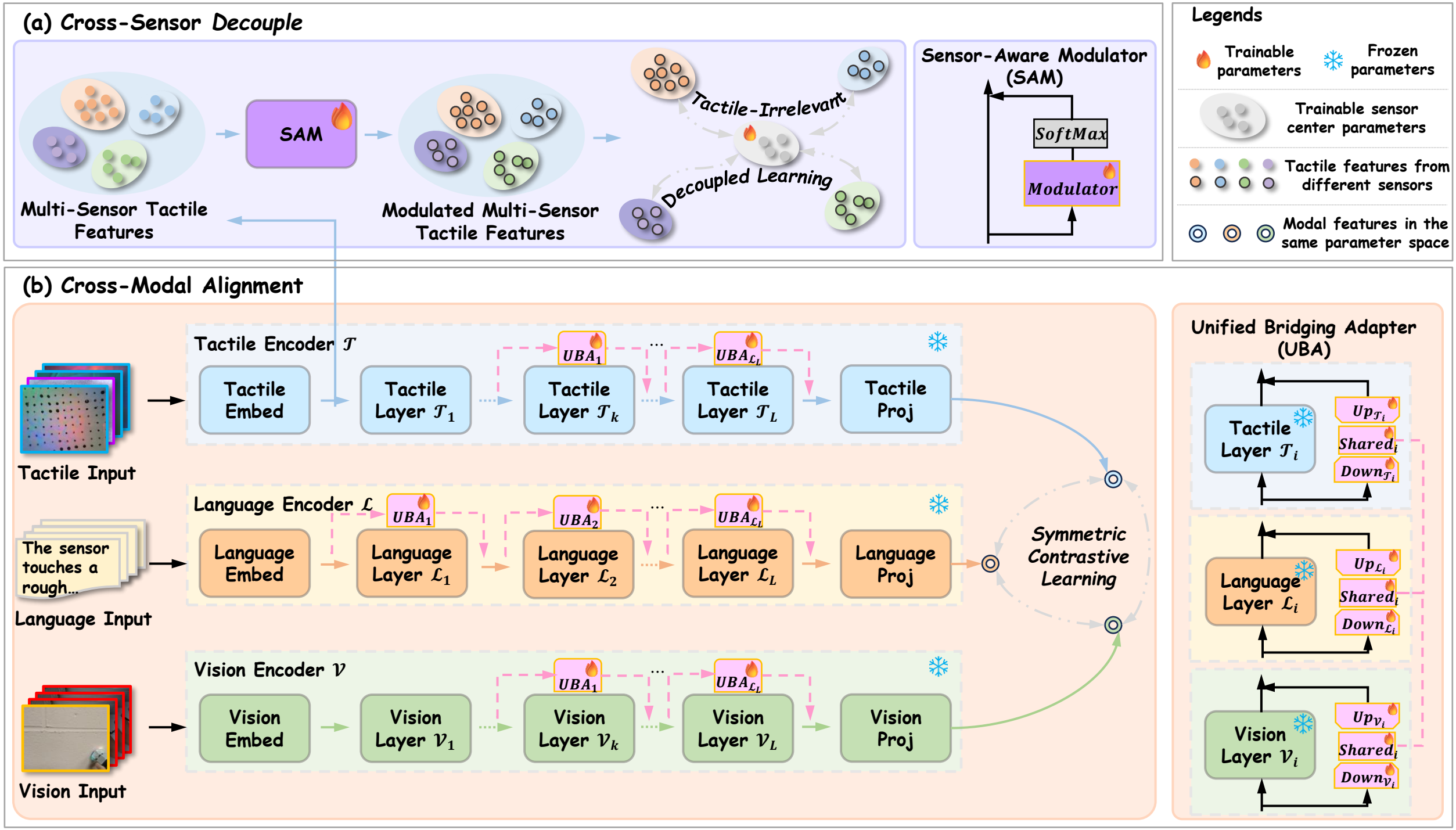
(a) Cross-Sensor Decouple
Multi-Sensor Tactile Features
SAM
Modulated Multi-Sensor Tactile Features
Tactile-Irrelevant
Decoupled Learning
Sensor-Aware Modulator (SAM)
SoftMax
Modulator
Legends
Trainable parameters
Frozen parameters
Trainable sensor center parameters
Tactile features from different sensors
Modal features in the same parameter space
(b) Cross-Modal Alignment
Tactile Input
Tactile Encoder $\mathcal{T}$
Tactile Embed
Tactile Layer $\mathcal{T}_1$
Tactile Layer $\mathcal{T}_k$
Tactile Layer $\mathcal{T}_L$
Tactile Proj
UBA_1
$UBA_{\mathcal{L}_L}$
The sensor touches a rough...
Language Input
Language Encoder $\mathcal{L}$
Language Embed
Language Layer $\mathcal{L}_1$
Language Layer $\mathcal{L}_2$
Language Layer $\mathcal{L}_L$
Language Proj
UBA_2
Vision Input
Vision Encoder $\mathcal{V}$
Vision Embed
Vision Layer $\mathcal{V}_1$
Vision Layer $\mathcal{V}_k$
Vision Layer $\mathcal{V}_L$
Vision Proj
Symmetric Contrastive Learning
Unified Bridging Adapter (UBA)
Tactile Layer $\mathcal{T}_i$
$Up_{\mathcal{T}_i}$
$Shared_i$
$Down_{\mathcal{T}_i}$
Language Layer $\mathcal{L}_i$
$Up_{\mathcal{L}_i}$
$Down_{\mathcal{L}_i}$
Vision Layer $\mathcal{V}_i$
$Up_{\mathcal{V}_i}$
$Down_{\mathcal{V}_i}$